Introduction to Probability

Lecture 10: Estimators (Part I)

Mateja Jamnik, Thomas Sauerwald

University of Cambridge, Department of Computer Science and Technology
email: {mateja.jamnik,thomas.sauerwald}@cl.cam.ac.uk

Easter 2025

UNIVERSITY OF CAMBRIDGE

Outline

Introduction

Introduction

> Setting: We can take random samples in the form of i.i.d. random variables $X_1, X_2, \ldots, X_n$ from an unknown distribution.

- Taking enough samples allows us to estimate the mean (WLLN, CLT)
- Using indicator variables, we can estimate $\mathbf{P}[X \leq a]$ for any $a \in \mathbb{R}$
  $\rightsquigarrow$ in principle we can reconstruct the entire distribution

Introduction

Setting: We can take random samples in the form of i.i.d. random variables $X_1, X_2, \ldots, X_n$ from an unknown distribution.

- Taking enough samples allows us to estimate the mean (WLLN, CLT)
- Using indicator variables, we can estimate $\mathbf{P}[X \leq a]$ for any $a \in \mathbb{R}$
  $\rightsquigarrow$ in principle we can reconstruct the entire distribution

???
- How can we directly estimate the variance or other parameters?
  $\rightsquigarrow$ estimator
- How can we measure the accuracy of an estimator?
  $\rightsquigarrow$ bias (this lecture) and mean squared error (next lecture)

Introduction

> Setting: We can take random samples in the form of i.i.d. random variables $X_1, X_2, \ldots, X_n$ from an unknown distribution.

- Taking enough samples allows us to estimate the mean (WLLN, CLT)
- Using indicator variables, we can estimate $\mathbf{P}[X \leq a]$ for any $a \in \mathbb{R}$
  $\rightsquigarrow$ in principle we can reconstruct the entire distribution

???
- How can we directly estimate the variance or other parameters?
  $\rightsquigarrow$ estimator
- How can we measure the accuracy of an estimator?
  $\rightsquigarrow$ bias (this lecture) and mean squared error (next lecture)

Physical Experiments:

Measurement $=$ Quantity of Interest $+$ Measurement Error

Introduction

> Setting: We can take random samples in the form of i.i.d. random variables $X_1, X_2, \ldots, X_n$ from an unknown distribution.

- Taking enough samples allows us to estimate the mean (WLLN, CLT)
- Using indicator variables, we can estimate $\mathbf{P}[X \leq a]$ for any $a \in \mathbb{R}$
  $\rightsquigarrow$ in principle we can reconstruct the entire distribution

???

- How can we directly estimate the variance or other parameters?
  $\rightsquigarrow$ estimator
- How can we measure the accuracy of an estimator?
  $\rightsquigarrow$ bias (this lecture) and mean squared error (next lecture)

expectation

Physical Experiments:

Measurement = Quantity of Interest + Measurement Error

Introduction

> Setting: We can take random samples in the form of i.i.d. random variables $X_1, X_2, \ldots, X_n$ from an unknown distribution.

- Taking enough samples allows us to estimate the mean (WLLN, CLT)
- Using indicator variables, we can estimate $\mathbf{P}[X \leq a]$ for any $a \in \mathbb{R}$
  $\rightsquigarrow$ in principle we can reconstruct the entire distribution

???

- How can we directly estimate the variance or other parameters?
  $\rightsquigarrow$ estimator
- How can we measure the accuracy of an estimator?
  $\rightsquigarrow$ bias (this lecture) and mean squared error (next lecture)

expectation   variance

Physical Experiments:

Measurement $=$ Quantity of Interest $+$ Measurement Error

Introduction

> Setting: We can take random samples in the form of i.i.d. random variables $X_1, X_2, \ldots, X_n$ from an unknown distribution.

- Taking enough samples allows us to estimate the mean (WLLN, CLT)
- Using indicator variables, we can estimate $\mathbf{P}[X \leq a]$ for any $a \in \mathbb{R}$
  $\rightsquigarrow$ in principle we can reconstruct the entire distribution

???

- How can we directly estimate the variance or other parameters?
  $\rightsquigarrow$ estimator
- How can we measure the accuracy of an estimator?
  $\rightsquigarrow$ bias (this lecture) and mean squared error (next lecture)

expectation   variance

Physical Experiments:

Measurement $=$ Quantity of Interest $+$ Measurement Error

Empirical Distribution Functions

Definition of Empirical Distribution Function (Empirical CDF)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. samples, and F be the corresponding distribution function. For any $a \in \mathbb{R}$, define

$$F_n(a) := \frac{\text{number of } X_i \in (-\infty, a]}{n}.$$

Empirical Distribution Functions

Definition of Empirical Distribution Function (Empirical CDF)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. samples, and F be the corresponding distribution function. For any $a \in \mathbb{R}$, define

$$F_n(a) := \frac{\text{number of } X_i \in (-\infty, a]}{n}.$$

Remark

The Weak Law of Large Numbers implies that for any $\epsilon > 0$ and $a \in \mathbb{R}$,

$$\lim_{n \to \infty} \mathbf{P}\left[\, |F_n(a) - F(a)| > \epsilon \,\right] = 0.$$

Empirical Distribution Functions

Definition of Empirical Distribution Function (Empirical CDF)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. samples, and F be the corresponding distribution function. For any $a \in \mathbb{R}$, define

$$F_n(a) := \frac{\text{number of } X_i \in (-\infty, a]}{n}.$$

Remark

The Weak Law of Large Numbers implies that for any $\epsilon > 0$ and $a \in \mathbb{R}$,

$$\lim_{n \to \infty} \mathbf{P}\left[\, |F_n(a) - F(a)| > \epsilon \,\right] = 0.$$

Thus by taking enough samples, we can estimate the entire distribution (including its expectation and variance).

Example 1

Consider throwing an unbiased dice 8 times, and let the realisation be:

$$(x_1, x_2, \ldots, x_8) = (4, 1, 4, 3, 1, 6, 4, 1).$$

What is the Empirical Distribution Function $F_8(a)$?
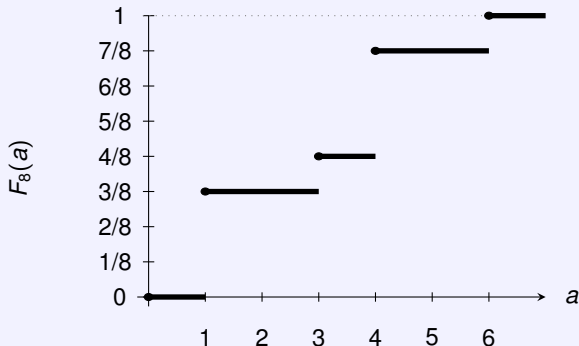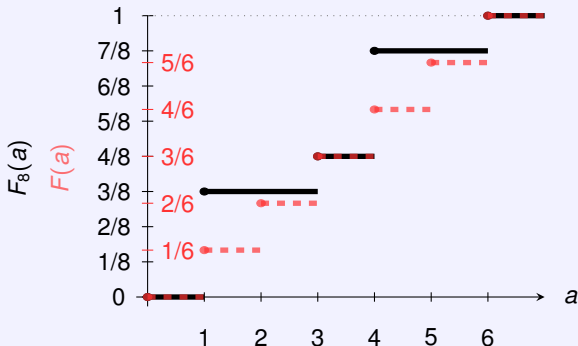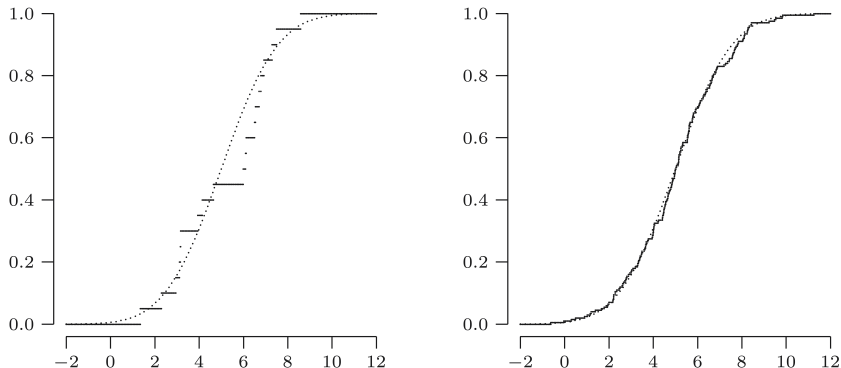
Answer

Empirical Distribution Functions (Example 1/2)

Consider throwing an unbiased dice 8 times, and let the realisation be:

$$(x_1, x_2, \ldots, x_8) = (4, 1, 4, 3, 1, 6, 4, 1).$$

What is the Empirical Distribution Function $F_8(a)$?

Answer

Example 1

Consider throwing an unbiased dice 8 times, and let the realisation be:

$$(x_1, x_2, \ldots, x_8) = (4, 1, 4, 3, 1, 6, 4, 1).$$

What is the Empirical Distribution Function $F_8(a)$?
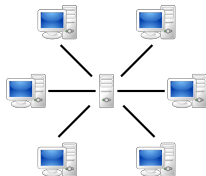
Answer

Empirical Distribution Functions (Example 2/2)

Source: Modern Introduction to Statistics

Figure: Empirical Distribution Functions of samples from a Normal Distribution $\mathcal{N}(5, 4)$ ($n = 20$ left, $n = 200$ right)

An Example of an Estimation Problem

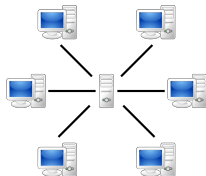Consider the packages arriving at a network server.

Source: Wikipedia

An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.
- We might be interested in:

Source: Wikipedia

An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.

- We might be interested in:
  1. number of packets that arrive within a "typical" minute
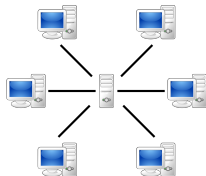

Source: Wikipedia

An Example of an Estimation Problem

— Scenario —

Consider the packages arriving at a network server.

- We might be interested in:
    1. number of packets that arrive within a "typical" minute

---

minutes

0   1   2   3   4   5   6

Source: Wikipedia

An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.
- We might be interested in:
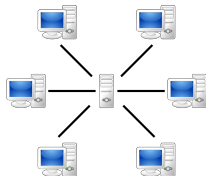 1. number of packets that arrive within a "typical" minute
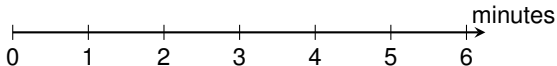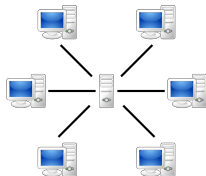

Source: Wikipedia

An Example of an Estimation Problem
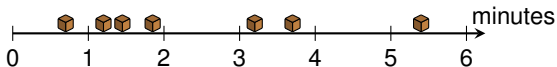

Scenario

Consider the packages arriving at a network server.

- We might be interested in:
  1. number of packets that arrive within a "typical" minute
  2. percentage of minutes during which no packets arrive

Source: Wikipedia

An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.
- We might be interested in:
    1. number of packets that arrive within a "typical" minute
    2. percentage of minutes during which no packets arrive
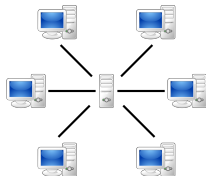

Source: Wikipedia

An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.

- We might be interested in:
  1. number of packets that arrive within a "typical" minute
  2. percentage of minutes during which no packets arrive

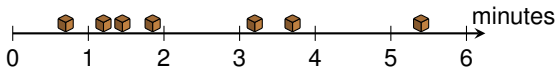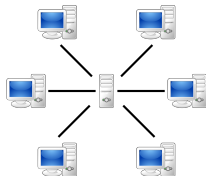Source: Wikipedia

An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.

- We might be interested in:
    1. number of packets that arrive within a "typical" minute
    2. percentage of minutes during which no packets arrive

- If arrivals occur at random time ⇝ number of arrivals during one minute follows a Poisson distribution with **unknown** parameter λ

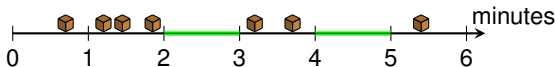Source: Wikipedia

An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.

- We might be interested in:
    1. number of packets that arrive within a "typical" minute
    2. percentage of minutes during which no packets arrive

- If arrivals occur at random time ⇝ number of arrivals during one minute follows a Poisson distribution with **unknown** parameter λ
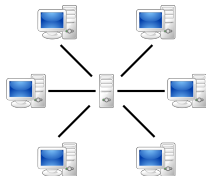

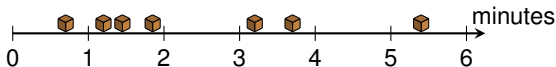Source: Wikipedia

An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.

- We might be interested in:
  1. number of packets that arrive within a "typical" minute
  2. percentage of minutes during which no packets arrive

- If arrivals occur at random time $\rightsquigarrow$ number of arrivals during one minute follows a Poisson distribution with **unknown** parameter λ

Source: Wikipedia

$X \sim Poi(\lambda)$

$$\boxed{\mathbf{P}\left[\,X = k\,\right] = e^{-\lambda} \cdot \frac{\lambda^k}{k!}}$$
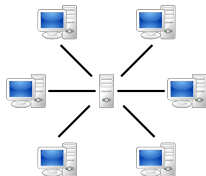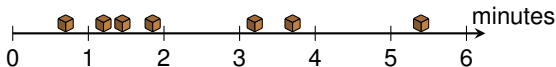
An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.

- We might be interested in:
  1. number of packets that arrive within a "typical" minute
  2. percentage of minutes during which no packets arrive

- If arrivals occur at random time $\rightsquigarrow$ number of arrivals during one minute follows a Poisson distribution with **unknown** parameter λ

Source: Wikipedia

$$X \sim Poi(\lambda)$$

$$\boxed{\mathbf{P}\left[\,X = k\,\right] = e^{-\lambda} \cdot \frac{\lambda^k}{k!}}$$

An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.

- We might be interested in:
    1. number of packets that arrive within a "typical" minute
    2. percentage of minutes during which no packets arrive

- If arrivals occur at random time $\rightsquigarrow$ number of arrivals during one minute follows a Poisson distribution with **unknown** parameter λ
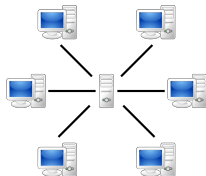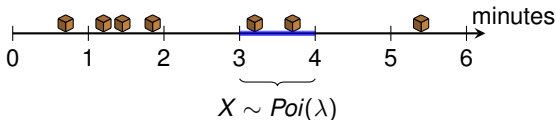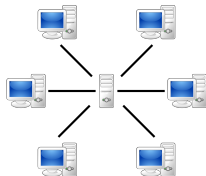

Source: Wikipedia

$\Delta \sim Exp(\lambda)$

minutes

$X \sim Poi(\lambda)$

$$\boxed{\mathbf{P}\,[\,X = k\,] = e^{-\lambda} \cdot \frac{\lambda^k}{k!}}$$
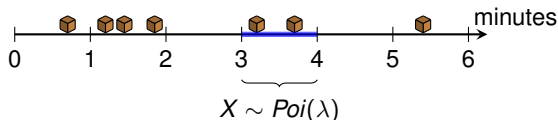
An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.
- We might be interested in:
    1. number of packets that arrive within a "typical" minute
    2. percentage of minutes during which no packets arrive

- If arrivals occur at random time ⇝ number of arrivals during one minute follows a Poisson distribution with **unknown** parameter λ

Waiting Time (Lecture 5, Slide 22)

$\Delta \sim Exp(\lambda)$

minutes

$X \sim Poi(\lambda)$

$$\mathbf{P}\left[\,X = k\,\right] = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

Source: Wikipedia

An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.

- We might be interested in:
    1. number of packets that arrive within a "typical" minute
    2. percentage of minutes during which no packets arrive

  > Estimator for λ

- If arrivals occur at random time $\rightsquigarrow$ number of arrivals during one minute follows a Poisson distribution with **unknown** parameter λ

> Waiting Time (Lecture 5, Slide 22)

$\Delta \sim Exp(\lambda)$

minutes
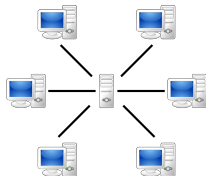
$X \sim Poi(\lambda)$

Source: Wikipedia

$$\mathbf{P}\,[\,X = k\,] = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$
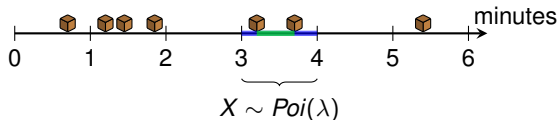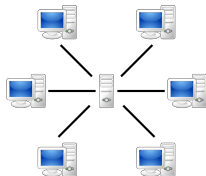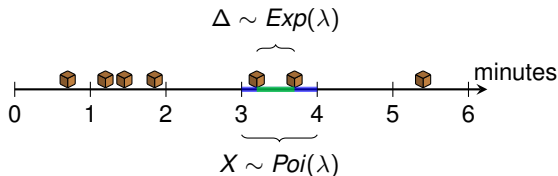
An Example of an Estimation Problem

Scenario

Consider the packages arriving at a network server.

- We might be interested in:
  1. number of packets that arrive within a "typical" minute — Estimator for λ
  2. percentage of minutes during which no packets arrive — Estimator for $e^{-\lambda}$

- If arrivals occur at random time ⤳ number of arrivals during one minute follows a Poisson distribution with **unknown** parameter λ

Waiting Time (Lecture 5, Slide 22)

$\Delta \sim Exp(\lambda)$

minutes

$X \sim Poi(\lambda)$

$$\mathbf{P}[X = k] = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

Source: Wikipedia

Estimator

Definition of Estimator

An estimate is a value t that only depends on the dataset $x_1, x_2, \ldots, x_n$, i.e.,

$$t = h(x_1, x_2, \ldots, x_n).$$

Estimator

Definition of Estimator

An estimate is a value t that only depends on the dataset $x_1, x_2, \ldots, x_n$, i.e.,

$$t = h(x_1, x_2, \ldots, x_n).$$

Then t is a realisation of the random variable

$$T = h(X_1, X_2, \ldots, X_n),$$

which is called estimator.

Estimator

Definition of Estimator

An estimate is a value t that only depends on the dataset $x_1, x_2, \ldots, x_n$, i.e.,

$$t = h(x_1, x_2, \ldots, x_n).$$

Then t is a realisation of the random variable

$$T = h(X_1, X_2, \ldots, X_n),$$

which is called estimator.

Questions:

Estimator

Definition of Estimator

An estimate is a value t that only depends on the dataset $x_1, x_2, \ldots, x_n$, i.e.,

$$t = h(x_1, x_2, \ldots, x_n).$$

Then t is a realisation of the random variable

$$T = h(X_1, X_2, \ldots, X_n),$$

which is called estimator.

Questions:
- What makes an estimator suitable? unbiased (later: mean squared error)
- Does an unbiased estimator always exist? How to compute it?
- If there are several unbiased estimators, which one to choose?

Outline

Introduction

Defining and Analysing Estimators

More Examples

Example: Arrival of Packets (1/3)

- Samples: Given $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim Pois(\lambda)$
- Meaning: X_i is the number of packets arriving in minute i

Example 2

Estimate λ by using the sample mean $\overline{X}_n$.

——————— Answer ———————

Example: Arrival of Packets (1/3)

- Samples: Given $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim Pois(\lambda)$
- Meaning: X_i is the number of packets arriving in minute i

Example 2

Estimate λ by using the sample mean $\overline{X}_n$.

———————————————————————— Answer ————————————————————————

We have
$$\overline{X}_n := \frac{X_1 + X_2 + \cdots + X_n}{n},$$

and $\mathbf{E}\left[\overline{X}_n\right] = \mathbf{E}[X_1] = \lambda$.

Example: Arrival of Packets (1/3)

- Samples: Given $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim Pois(\lambda)$
- Meaning: X_i is the number of packets arriving in minute i

Example 2

Estimate λ by using the sample mean $\overline{X}_n$.

—————————————————————— Answer ——————————————————————

We have

$$\overline{X}_n := \frac{X_1 + X_2 + \cdots + X_n}{n},$$

and $\mathbf{E}\left[\overline{X}_n\right] = \mathbf{E}[X_1] = \lambda$. This suggests the estimator:

$$h(X_1, X_2, \ldots, X_n) := \overline{X}_n.$$

- Samples: Given $X_1, X_2, \ldots, X_n$ i.i.d., $X_i \sim Pois(\lambda)$
- Meaning: X_i is the number of packets arriving in minute i

Example 2

Estimate λ by using the sample mean $\overline{X}_n$.

———————————————————————————————— Answer ————————————————————————————————

We have

$$\overline{X}_n := \frac{X_1 + X_2 + \cdots + X_n}{n},$$

and $\mathbf{E}\left[\overline{X}_n\right] = \mathbf{E}[X_1] = \lambda$. This suggests the estimator:

$$h(X_1, X_2, \ldots, X_n) := \overline{X}_n.$$

Applying the Weak Law of Large Numbers:

$$\lim_{n \to \infty} \mathbf{P}\left[\left|\overline{X}_n - \lambda\right| > \epsilon\right] = 0 \qquad \text{for any } \epsilon > 0.$$

Example 3a

Define an estimator h_1 for the probability of zero arrivals, $e^{-\lambda}$.

Answer

Example 3a

Define an estimator h_1 for the probability of zero arrivals, $e^{-\lambda}$.

———————— Answer ————————

Let $X_1, X_2, \ldots, X_n$ be the n samples. Let

$$Y_i := \mathbf{1}_{X_i = 0}.$$

Example 3a

Define an estimator h_1 for the probability of zero arrivals, $e^{-\lambda}$.

Answer

Let $X_1, X_2, \ldots, X_n$ be the n samples. Let

$$Y_i := \mathbf{1}_{X_i = 0}.$$

Then

$$\mathbf{E}[Y_i] = \mathbf{P}[X_i = 0] = e^{-\lambda},$$

Example 3a

Define an estimator h_1 for the probability of zero arrivals, $e^{-\lambda}$.

Answer

Let $X_1, X_2, \ldots, X_n$ be the n samples. Let

$$Y_i := \mathbf{1}_{X_i=0}.$$

Then

$$\mathbf{E}\,[\,Y_i\,] = \mathbf{P}\,[\,X_i = 0\,] = e^{-\lambda},$$

and thus we can define an estimator by

$$h_1(X_1, X_2, \ldots, X_n) := \frac{Y_1 + Y_2 + \cdots + Y_n}{n} = \overline{Y_n}.$$

- Suppose we get the samples $(x_1, x_2, x_3) = (50, 100, 0)$

- Suppose we get the samples $(x_1, x_2, x_3) = (50, 100, 0)$
- Then $(y_1, y_2, y_3) = (0, 0, 1)$, and $h_1(x_1, x_2, x_3) = \frac{1}{3}$

- Suppose we get the samples $(x_1, x_2, x_3) = (50, 100, 0)$
- Then $(y_1, y_2, y_3) = (0, 0, 1)$, and $h_1(x_1, x_2, x_3) = \frac{1}{3}$
- This seems too large! Also note that for the samples $(x_1, x_2, x_3) = (1, 1, 0)$, our estimator would give the same estimate

Example: Arrival of Packets (3/3)

- Suppose we get the samples $(x_1, x_2, x_3) = (50, 100, 0)$
- Then $(y_1, y_2, y_3) = (0, 0, 1)$, and $h_1(x_1, x_2, x_3) = \frac{1}{3}$
- This seems too large! Also note that for the samples $(x_1, x_2, x_3) = (1, 1, 0)$, our estimator would give the same estimate

<div>

Example 3b

Define an estimator h_2 for $e^{-\lambda}$ based on $\overline{X}_n$.

———————————————————— Answer ————————————————————

</div>

Example: Arrival of Packets (3/3)

- Suppose we get the samples $(x_1, x_2, x_3) = (50, 100, 0)$
- Then $(y_1, y_2, y_3) = (0, 0, 1)$, and $h_1(x_1, x_2, x_3) = \frac{1}{3}$
- This seems too large! Also note that for the samples
  $(x_1, x_2, x_3) = (1, 1, 0)$, our estimator would give the same estimate

---

Example 3b

Define an estimator h_2 for $e^{-\lambda}$ based on $\overline{X}_n$.

Answer

We saw that $\overline{X}_n = \frac{\sum_{i=1}^n X_i}{n}$ satisfies $\mathbf{E}\left[\overline{X}_n \right] = \mathbf{E}[X_1] = \lambda$.

Recall by the Weak Law of Large Numbers:

Example: Arrival of Packets (3/3)

- Suppose we get the samples $(x_1, x_2, x_3) = (50, 100, 0)$
- Then $(y_1, y_2, y_3) = (0, 0, 1)$, and $h_1(x_1, x_2, x_3) = \frac{1}{3}$
- This seems too large! Also note that for the samples $(x_1, x_2, x_3) = (1, 1, 0)$, our estimator would give the same estimate

---

Example 3b

Define an estimator h_2 for $e^{-\lambda}$ based on $\overline{X}_n$.

———————————————————————— Answer ————————————————————————

We saw that $\overline{X}_n = \frac{\sum_{i=1}^n X_i}{n}$ satisfies $\mathbf{E}\left[\overline{X}_n\right] = \mathbf{E}[X_1] = \lambda$.

Recall by the Weak Law of Large Numbers:

$$\lim_{n\to\infty} \mathbf{P}\left[\left|\overline{X}_n - \lambda\right| > \epsilon\right] = 0 \qquad \text{for any } \epsilon > 0.$$

Example: Arrival of Packets (3/3)

- Suppose we get the samples $(x_1, x_2, x_3) = (50, 100, 0)$
- Then $(y_1, y_2, y_3) = (0, 0, 1)$, and $h_1(x_1, x_2, x_3) = \frac{1}{3}$
- This seems too large! Also note that for the samples
  $(x_1, x_2, x_3) = (1, 1, 0)$, our estimator would give the same estimate

---

Example 3b

Define an estimator h_2 for $e^{-\lambda}$ based on $\overline{X}_n$.

─────────────────────── Answer ───────────────────────

We saw that $\overline{X}_n = \frac{\sum_{i=1}^n X_i}{n}$ satisfies $\mathbf{E}\left[\overline{X}_n\right] = \mathbf{E}[X_1] = \lambda$.

Recall by the Weak Law of Large Numbers:

$$\lim_{n \to \infty} \mathbf{P}\left[\left|\overline{X}_n - \lambda\right| > \epsilon\right] = 0 \qquad \text{for any } \epsilon > 0.$$

This suggests to estimate $e^{-\lambda}$ by $e^{-\overline{X}_n}$. Hence our estimator is

$$h_2(X_1, X_2, \dots, X_n) := e^{-\overline{X}_n}.$$

- Suppose we have $n = 30$ and we want to estimate $e^{-\lambda}$
- Consider the two estimators $h_1(X_1, \ldots, X_n)$ and $h_2(X_1, \ldots, X_n)$.

- Suppose we have $n = 30$ and we want to estimate $e^{-\lambda}$
- Consider the two estimators $h_1(X_1, \ldots, X_n)$ and $h_2(X_1, \ldots, X_n)$.

How **good** are these two estimators?

- Suppose we have $n = 30$ and we want to estimate $e^{-\lambda}$
- Consider the two estimators $h_1(X_1, \ldots, X_n)$ and $h_2(X_1, \ldots, X_n)$.

How **good** are these two estimators?

$\Rightarrow$ The first estimator can only attain values $0, \frac{1}{30}, \frac{2}{30}, \ldots, 1$

$\Rightarrow$ The second estimator can only attain values $1, e^{-1/30}, e^{-2/30}, \ldots$

Behaviour of the Estimators

- Suppose we have $n = 30$ and we want to estimate $e^{-\lambda}$
- Consider the two estimators $h_1(X_1, \ldots, X_n)$ and $h_2(X_1, \ldots, X_n)$.

How **good** are these two estimators?

$\Rightarrow$ The first estimator can only attain values $0, \frac{1}{30}, \frac{2}{30}, \ldots, 1$

$\Rightarrow$ The second estimator can only attain values $1, e^{-1/30}, e^{-2/30}, \ldots$

For most values of λ, both estimators will never return the exact value of $e^{-\lambda}$ on the basis of 30 observations.

Simulation of the two Estimators

- The unknown parameter is $p = e^{-\lambda} = 0.1$ (i.e., $\lambda = \ln 10 \approx 2.30\ldots$)

Simulation of the two Estimators

- The unknown parameter is $p = e^{-\lambda} = 0.1$ (i.e., $\lambda = \ln 10 \approx 2.30\ldots$)
- We consider $n = 30$ minutes and compute h_1 and h_2
- We repeat this 500 times and draw a frequency histogram
  ($h_1 = \overline{Y}_n$ left, $h_2 = e^{-\overline{X}_n}$ right)

Simulation of the two Estimators

- The unknown parameter is $p = e^{-\lambda} = 0.1$ (i.e., $\lambda = \ln 10 \approx 2.30\ldots$)
- We consider $n = 30$ minutes and compute h_1 and h_2
- We repeat this 500 times and draw a frequency histogram
  ($h_1 = \overline{Y}_n$ left, $h_2 = e^{-\overline{X}_n}$ right)
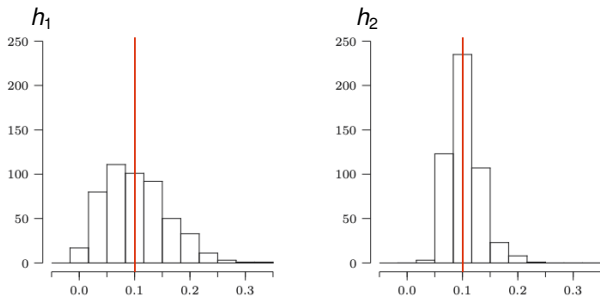
Source: Modern Introduction to Statistics

Simulation of the two Estimators

- The unknown parameter is $p = e^{-\lambda} = 0.1$ (i.e., $\lambda = \ln 10 \approx 2.30\ldots$)
- We consider $n = 30$ minutes and compute h_1 and h_2
- We repeat this 500 times and draw a frequency histogram
  ($h_1 = \overline{Y}_n$ left, $h_2 = e^{-\overline{X}_n}$ right)
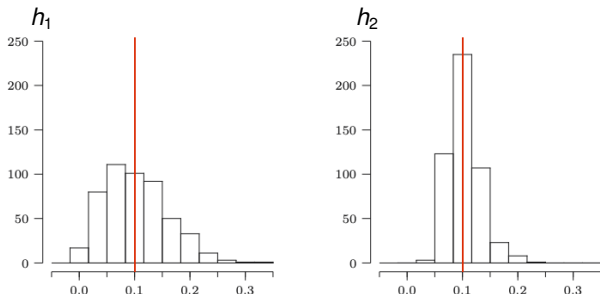
Source: Modern Introduction to Statistics

> Both estimators concentrate around the true value 0.1, but
> the second estimator appears to be more concentrated.

Unbiased Estimators and Bias

— Definition —

An estimator T is called an unbiased estimator for the parameter θ if

$$\mathbf{E}\,[\,T\,] = \theta,$$

irrespective of the value θ.

Unbiased Estimators and Bias

Definition

An estimator *T* is called an unbiased estimator for the parameter θ if

$$\mathbf{E}[\,T\,] = \theta,$$

irrespective of the value θ. The bias is defined as

$$\mathbf{E}[\,T\,] - \theta = \mathbf{E}[\,T - \theta\,].$$

Source: Edwin Leuven (Point Estimation)

Definition

An estimator *T* is called an unbiased estimator for the parameter θ if

$$\mathbf{E}\,[\,T\,] = \theta,$$

irrespective of the value θ. The bias is defined as

$$\mathbf{E}\,[\,T\,] - \theta = \mathbf{E}\,[\,T - \theta\,].$$

Unbiased Estimators and Bias

Definition

An estimator T is called an unbiased estimator for the parameter θ if

$$\mathbf{E}[\,T\,] = \theta,$$
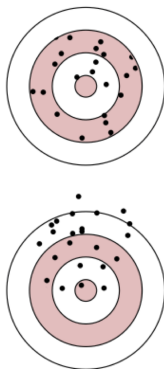
irrespective of the value θ. The bias is defined as

$$\mathbf{E}[\,T\,] - \theta = \mathbf{E}[\,T - \theta\,].$$
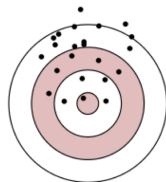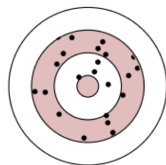
Source: Edwin Leuven (Point Estimation)

Which of the two estimators h_1, h_2 are unbiased?

Example 4a

Is $h_1(X_1, X_2, \ldots, X_n) = \frac{Y_1 + Y_2 + \cdots + Y_n}{n}$ an unbiased estimator for $e^{-\lambda}$?

Answer

Example 4a

Is $h_1(X_1, X_2, \ldots, X_n) = \frac{Y_1 + Y_2 + \cdots + Y_n}{n}$ an unbiased estimator for $e^{-\lambda}$?

—————————————— Answer ——————————————

Recall we defined $Y_i := \mathbf{1}_{X_i=0}$.

Analysis of the Bias of the First Estimator

Example 4a

Is $h_1(X_1, X_2, \ldots, X_n) = \frac{Y_1 + Y_2 + \cdots + Y_n}{n}$ an unbiased estimator for $e^{-\lambda}$?

Answer

Recall we defined $Y_i := \mathbf{1}_{X_i = 0}$. **Yes**, because:

$$\mathbf{E}\left[\, h_1(X_1, X_2, \ldots, X_n)\,\right]$$

Example 4a

Is $h_1(X_1, X_2, \ldots, X_n) = \frac{Y_1 + Y_2 + \cdots + Y_n}{n}$ an unbiased estimator for $e^{-\lambda}$?

———————————————— Answer ————————————————

Recall we defined $Y_i := \mathbf{1}_{X_i=0}$. **Yes**, because:

$$\mathbf{E}[h_1(X_1, X_2, \ldots, X_n)] = \frac{n \cdot \mathbf{E}[Y_1]}{n}$$

Example 4a

Is $h_1(X_1, X_2, \ldots, X_n) = \frac{Y_1 + Y_2 + \cdots + Y_n}{n}$ an unbiased estimator for $e^{-\lambda}$?

Answer

Recall we defined $Y_i := \mathbf{1}_{X_i = 0}$. **Yes**, because:

$$\mathbf{E}[h_1(X_1, X_2, \ldots, X_n)] = \frac{n \cdot \mathbf{E}[Y_1]}{n}$$
$$= \mathbf{P}[X_1 = 0]$$

Analysis of the Bias of the First Estimator

Example 4a

Is $h_1(X_1, X_2, \ldots, X_n) = \frac{Y_1 + Y_2 + \cdots + Y_n}{n}$ an unbiased estimator for $e^{-\lambda}$?

_____ Answer _____

Recall we defined $Y_i := \mathbf{1}_{X_i = 0}$. **Yes**, because:

$$\begin{aligned}
\mathbf{E}[h_1(X_1, X_2, \ldots, X_n)] &= \frac{n \cdot \mathbf{E}[Y_1]}{n} \\
&= \mathbf{P}[X_1 = 0] \\
&= e^{-\lambda}.
\end{aligned}$$

Bias of the Second Estimator (and Jensen's Inequality)

Example 4b

Is $h_2(X_1, X_2, \ldots, X_n) = e^{-\overline{X}_n}$ an unbiased estimator for $e^{-\lambda}$?

Answer

Bias of the Second Estimator (and Jensen's Inequality)

Example 4b

Is $h_2(X_1, X_2, \ldots, X_n) = e^{-\overline{X}_n}$ an unbiased estimator for $e^{-\lambda}$?

—————— Answer ——————

No! (recall: $\mathbf{E}\left[X^2\right] \geq \mathbf{E}\left[X\right]^2$)

Bias of the Second Estimator (and Jensen's Inequality)

Example 4b

Is $h_2(X_1, X_2, \ldots, X_n) = e^{-\overline{X}_n}$ an unbiased estimator for $e^{-\lambda}$?

Answer

No! (recall: $\mathbf{E}\left[X^2\right] \geq \mathbf{E}\left[X\right]^2$)

Jensen's Inequality

For any random variable X, and any convex function $g : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbf{E}\left[g(X)\right] \geq g(\mathbf{E}\left[X\right]).$$

If g is strictly convex and X is not constant, then the inequality is strict.

Bias of the Second Estimator (and Jensen's Inequality)

Example 4b

Is $h_2(X_1, X_2, \ldots, X_n) = e^{-\overline{X}_n}$ an unbiased estimator for $e^{-\lambda}$?

Answer

No! (recall: $\mathbf{E}\left[X^2\right] \geq \mathbf{E}\left[X\right]^2$)

$$\lambda g(a) + (1 - \lambda)g(b) \geq g(\lambda a + (1 - \lambda)b)$$

Jensen's Inequality

For any random variable X, and any convex function $g : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbf{E}\left[g(X)\right] \geq g(\mathbf{E}\left[X\right]).$$

If g is strictly convex and X is not constant, then the inequality is strict.

Bias of the Second Estimator (and Jensen's Inequality)

Example 4b

Is $h_2(X_1, X_2, \ldots, X_n) = e^{-\overline{X}_n}$ an unbiased estimator for $e^{-\lambda}$?

Answer

No! (recall: $\mathbf{E}\left[X^2\right] \geq \mathbf{E}\left[X\right]^2$)

- We have

$$\mathbf{E}\left[e^{-\overline{X}_n}\right] > e^{-\mathbf{E}\left[\overline{X}_n\right]} = e^{-\lambda}$$

$$\lambda g(a) + (1-\lambda)g(b) \geq g(\lambda a + (1-\lambda)b)$$

Jensen's Inequality

For any random variable X, and any convex function $g : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbf{E}\left[g(X)\right] \geq g(\mathbf{E}\left[X\right]).$$

If g is strictly convex and X is not constant, then the inequality is strict.

Bias of the Second Estimator (and Jensen's Inequality)

> **Example 4b**
>
> Is $h_2(X_1, X_2, \ldots, X_n) = e^{-\overline{X}_n}$ an unbiased estimator for $e^{-\lambda}$?
>
> ———————————————— Answer ————————————————
>
> **No! (recall: $\mathbf{E}\left[X^2\right] \geq \mathbf{E}\left[X\right]^2$)**
>
> - We have
>
> $$\mathbf{E}\left[e^{-\overline{X}_n}\right] > e^{-\mathbf{E}\left[\overline{X}_n\right]} = e^{-\lambda}$$
>
> - This follows by Jensen's inequality, and the inequality is strict since $g : z \mapsto e^{-z}$ is strictly convex and $\overline{X}_n$ is not constant.

$$\boxed{\lambda g(a) + (1 - \lambda)g(b) \ \geq \ g(\lambda a + (1 - \lambda)b)}$$

———— **Jensen's Inequality** ————

For any random variable X, and any convex function $g : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbf{E}\left[g(X)\right] \geq g(\mathbf{E}\left[X\right]).$$

If g is strictly convex and X is not constant, then the inequality is strict.

Bias of the Second Estimator (and Jensen's Inequality)

Example 4b

Is $h_2(X_1, X_2, \ldots, X_n) = e^{-\overline{X}_n}$ an unbiased estimator for $e^{-\lambda}$?

Answer

No! (recall: $\mathbf{E}\left[X^2\right] \geq \mathbf{E}\left[X\right]^2$)

- We have

$$\mathbf{E}\left[e^{-\overline{X}_n}\right] > e^{-\mathbf{E}\left[\overline{X}_n\right]} = e^{-\lambda}$$

- This follows by Jensen's inequality, and the inequality is strict since $g : z \mapsto e^{-z}$ is strictly convex and $\overline{X}_n$ is not constant.

- Thus $h_2(X_1, X_2, \ldots, X_n)$ is not unbiased – it has positive bias.

$$\lambda g(a) + (1 - \lambda)g(b) \geq g(\lambda a + (1 - \lambda)b)$$

Jensen's Inequality

For any random variable X, and any convex function $g : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbf{E}\left[g(X)\right] \geq g(\mathbf{E}\left[X\right]).$$

If g is strictly convex and X is not constant, then the inequality is strict.

Asymptotic Bias of the Second Estimator (non-examinable)

Example 4c

$\mathbf{E}\left[\, h_2(X_1, \dots, X_n)\,\right] \overset{n \to \infty}{\longrightarrow} e^{-\lambda}$ (hence it is asymptotically unbiased).

——————————————— Answer ———————————————

- Recall $h_2(X_1, \dots, X_n) = e^{-\overline{X}_n}$. For any $0 \le k \le n$,

$$\mathbf{P}\left[\, h_2(X_1, \dots, X_n) = e^{-k/n}\,\right] = \mathbf{P}\left[\sum_{i=1}^{n} X_i = k\right] = \mathbf{P}\left[\, Z = k\,\right],$$

where $Z \sim Pois(n \cdot \lambda)$ (since $Pois(\lambda_1) + Pois(\lambda_2) = Pois(\lambda_1 + \lambda_2)$)

$$\Rightarrow \qquad \mathbf{P}\left[\, h_2(X_1, \dots, X_n) = e^{-k/n}\,\right] = \frac{e^{-n\lambda} \cdot (n\lambda)^k}{k!}$$

$$\Rightarrow \qquad \mathbf{E}\left[\, h_2(X_1, \dots, X_n)\,\right] = \sum_{k=0}^{\infty} e^{-n\lambda} \cdot \frac{(n\lambda^k)}{k!} \cdot e^{-k/n}$$

$\boxed{\text{By LOTUS}}$

$$= e^{-n\lambda} \cdot e^{n\lambda e^{-1/n}} \sum_{k=0}^{\infty} e^{-n\lambda e^{-1/n}} \cdot \frac{(n\lambda e^{-1/n})^k}{k!}$$

$$= e^{-n\lambda \cdot (1 - e^{-1/n})} \cdot 1$$

$\boxed{\text{since } e^x = 1 + x + O(x^2) \text{ for small } x}$ $\overset{n \to \infty}{\approx} e^{-n\lambda \cdot (1 - 1 + 1/n + O(1/n^2))} = e^{-\lambda + O(\lambda/n)}.$

$\boxed{\text{Hence in the limit, the positive bias of } h_2 \text{ diminishes.}}$

Outline

Introduction

Defining and Analysing Estimators

More Examples

Unbiased Estimator for Expectation and Variance

Unbiased Estimators for Expectation and Variance

Let $X_1, X_2, \ldots, X_n$ be identically distributed samples from a distribution with finite expectation μ and finite variance σ^2.

- Then
$$\overline{X}_n := \frac{X_1 + X_2 + \cdots + X_n}{n}$$

  is an unbiased estimator for μ.

- Furthermore, for $n \geq 2$,
$$S_n = S_n(X_1, \ldots, X_n) := \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(X_i - \overline{X}_n \right)^2$$

  is an unbiased estimator for σ^2.

Example 5

We need to prove: $\mathbf{E}\,[\,S_n\,] = \sigma^2$.

———————————————— Answer ————————————————

Multiplying by $n - 1$ yields:

$$(n - 1) \cdot S_n = \sum_{i=1}^{n} \left(X_i - \overline{X}_n \right)^2$$

$$= \sum_{i=1}^{n} \left(X_i - \mu + \mu - \overline{X}_n \right)^2$$

$$= \sum_{i=1}^{n} (X_i - \mu)^2 + \sum_{i=1}^{n} \left(\overline{X}_n - \mu \right)^2 - 2 \sum_{i=1}^{n} (X_i - \mu) \left(\overline{X}_n - \mu \right)$$

$$= \sum_{i=1}^{n} (X_i - \mu)^2 + n \left(\overline{X}_n - \mu \right)^2 - 2 \left(\overline{X}_n - \mu \right) \cdot n \cdot \left(\overline{X}_n - \mu \right)$$

$$= \sum_{i=1}^{n} (X_i - \mu)^2 - n \left(\overline{X}_n - \mu \right)^2 .$$

Let us now take expectations:

> **By Lec. 8, Slide 21: $\mathbf{E}\left[(\overline{X}_n - \mu)^2 \right] = \mathbf{V}\left[\overline{X}_n \right] = \sigma^2/n$**

$$(n - 1) \cdot \mathbf{E}\,[\,S_n\,] = \sum_{i=1}^{n} \mathbf{E}\left[(X_i - \mu)^2 \right] - n \cdot \mathbf{E}\left[\left(\overline{X}_n - \mu \right)^2 \right]$$

$$= n \cdot \sigma^2 - n \cdot \sigma^2/n$$

$$= (n - 1) \cdot \sigma^2 .$$

An Unbiased Estimator may not always exist

Example 6

Suppose that we have one sample $X \sim Bin(n, p)$, where $0 < p < 1$ is unknown but n is known. Prove there is no unbiased estimator for $1/p$.

———————————— Answer ————————————

- First a simpler proof which exploits that p might be arbitrarily small
- Intuition: By making p smaller and smaller, we force $\max_{0 \leq k \leq n} T(k)$, $k \in \{0, 1, \ldots, n\}$ to become bigger and bigger
- Formal Argument:
  - Fix any estimator $T(X)$
  - Define $M := \max_{0 \leq k \leq n} T(k)$. Then,

$$\mathbf{E}\,[\,T(X)\,] = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \cdot T(k)$$

$$\leq M \cdot \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = M.$$

  - Hence this estimator does not work for $p < \frac{1}{M}$, since then $\mathbf{E}\,[\,T(X)\,] \leq M < \frac{1}{p}$ (negative bias!)
- The next proof will work even if $p \in [a, b]$ for $0 < a < b \leq 1$.

An Unbiased Estimator may not always exist (cntd. - non-examinable)

Example 6 (cntd.)

Suppose that we have one sample $X \sim Bin(n, p)$, where $0 < p < 1$ is unknown but n is known. Prove there is no unbiased estimator for $1/p$.

Answer

- Suppose there exists an unbiased estimator with $\mathbf{E}[T(X)] = 1/p$.
- Then

$$1 = p \cdot \mathbf{E}[T(X)]$$

$$= p \cdot \sum_{k=0}^{n} \mathbf{P}[X = k] \cdot T(k)$$

$$= p \cdot \sum_{k=0}^{n} \binom{n}{k} p^k \cdot (1 - p)^{n-k} \cdot T(k)$$

- Last term is a polynomial of degree $n + 1$ with constant term zero
  $\Rightarrow p \cdot \mathbf{E}[T(X)] - 1$ is a (non-zero) polynomial of degree $\leq n + 1$
  $\Rightarrow$ this polynomial has at most $n + 1$ roots
  $\Rightarrow \mathbf{E}[T(X)]$ can be equal to $1/p$ for at most $n + 1$ values of p, and thus cannot be an unbiased.